



The Human Genome Project

Dr. Manfred Zorn

**April 12, 2001
School of Public Health
UC Berkeley**



Lawrence Berkeley National Laboratory



- **Founded in 1931 by Ernest O. Lawrence**
- **Best known for Particle Physics, found a dozen new transuranic elements: Bk, Cf, Am, Lw, Pu, ..., Sg**
- **About 4000 people, 800 students, 2000 visitors**
- **National User Facilities:**
 - **Advanced Light Source**
 - **NERSC Supercomputing Center**

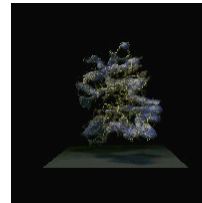
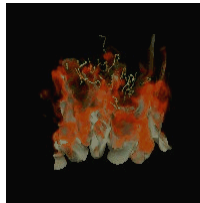
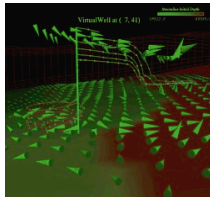
UCB Seminar



NERSC - Overview



- **the Department of Energy, Office of Science, supercomputing facility**
- **unclassified, open facility; serving >2000 users in all DOE mission relevant basic science disciplines**
- **25th anniversary in 1999**



UCB Seminar



Center for Bioinformatics and Computational Genomics



- **Research**
 - **Special Analysis Tools: Fold Prediction, Phylogeny, genome comparisons**
 - **Compute-intensive Algorithms: clustering, phylogeny**
- **Development and Support**
 - **Large-scale Genome Annotation**
 - **Wet lab support for Biologists**
- **Public Service**
 - **Public databases**
 - **Education and Outreach, Standards**

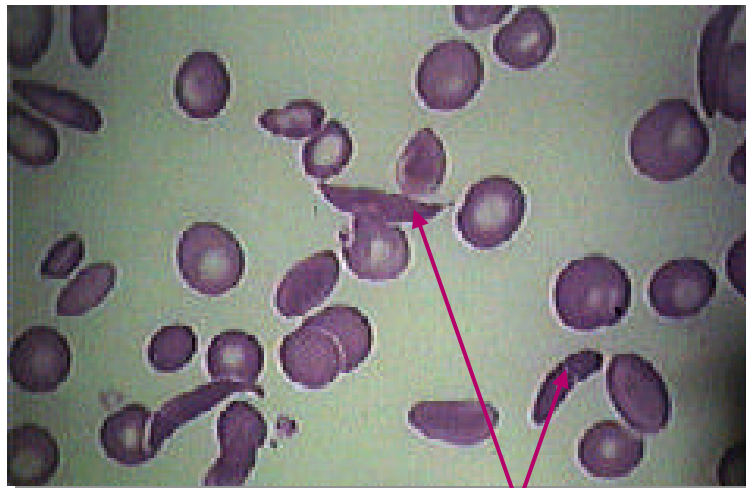
UCB Seminar

3 Questions

- **What is the genome?**
- **Why is it important?**
- **What are the consequences?**

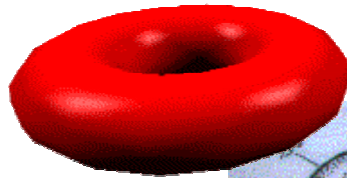
UCB Seminar

Diagnosis - Blood Smear

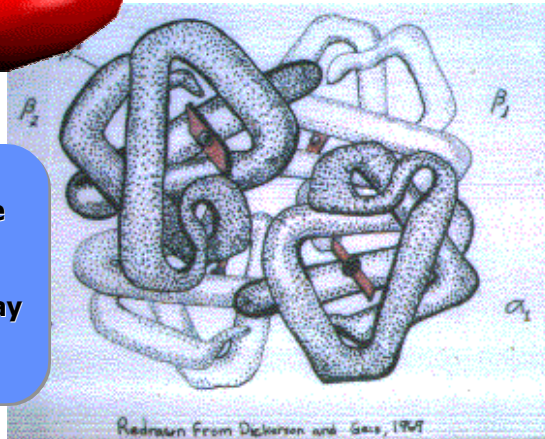


Sickle red cells

UCB Seminar



Hemoglobin is the main chemical in the red blood cell that does all of the work carrying oxygen away from the lungs and carbon dioxide back



UCB Seminar

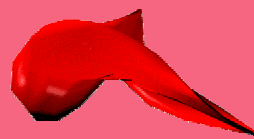
Normal

- disc-Shaped
- soft (like a bag of jelly)
- easily flow through small blood vessels
- lives for 120 days

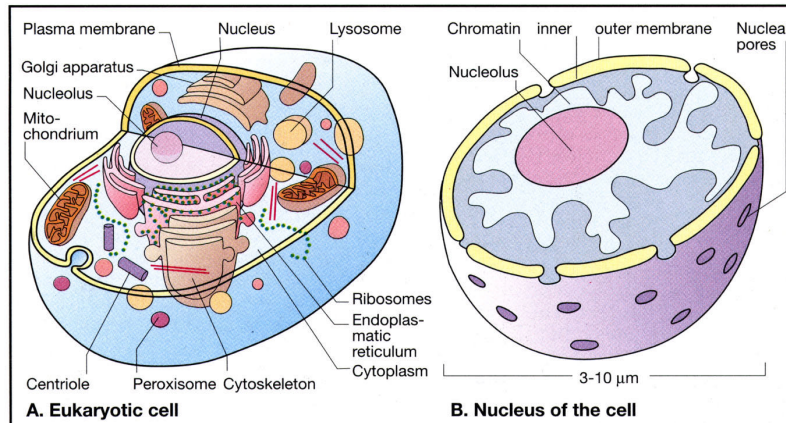


Sickle

- sickle-Shaped
- hard (like a piece of wood)
- often get stuck in small blood vessels
- lives for 20 days or less



UCB Seminar



ZBD9806-01631.TIF

UCB Seminar



Truth and Conventional Wisdom in Biology

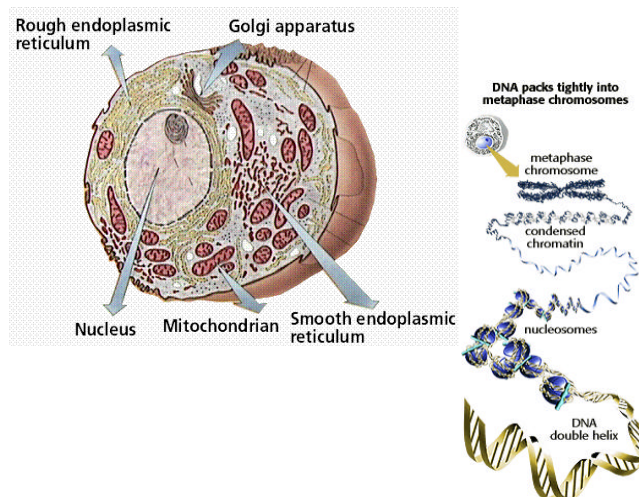
- *Biologists dislike generalizations*
- *The truth in biology is always more complex than the statement about it*
- *It is hard to distinguish between fact and fashion in biology*

Chromosomes



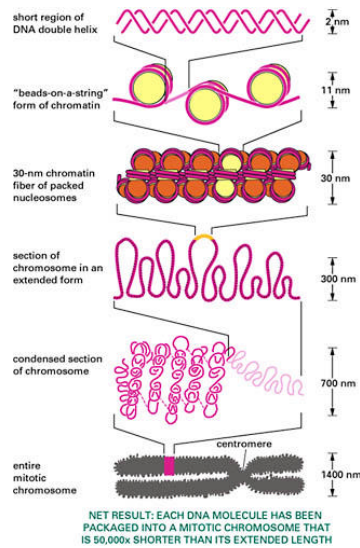
UCB Seminar

Basic Biology



UCB Seminar

Scale



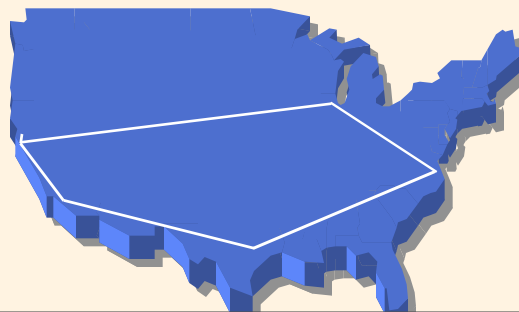
UCB Seminar

The Human Genome

- **24 Chromosomes**
 - ✓ 1 – 22, X, Y
 - ✓ 23 pairs
- **1 Mitochondrial Genome**
- **3 Billion Base Pairs**
- **~30,000 Genes**

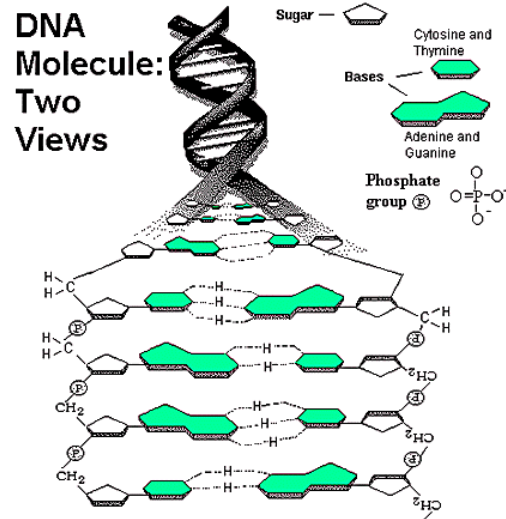
UCB Seminar

3×10^9 seconds = 95 years



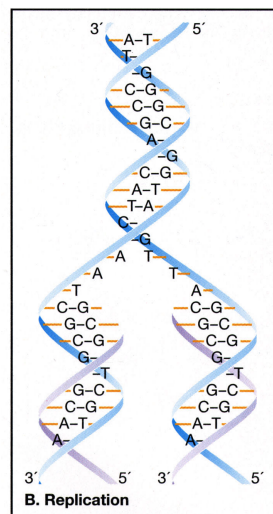
Typed in 10-pitch font, one human sequence would stretch for more than 5,000 miles. Digitally formatted, it could be stored on one CD-ROM. Biologically encoded, it fits easily within a single cell.

DNA - Two Views



UCB Seminar

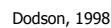
Replication



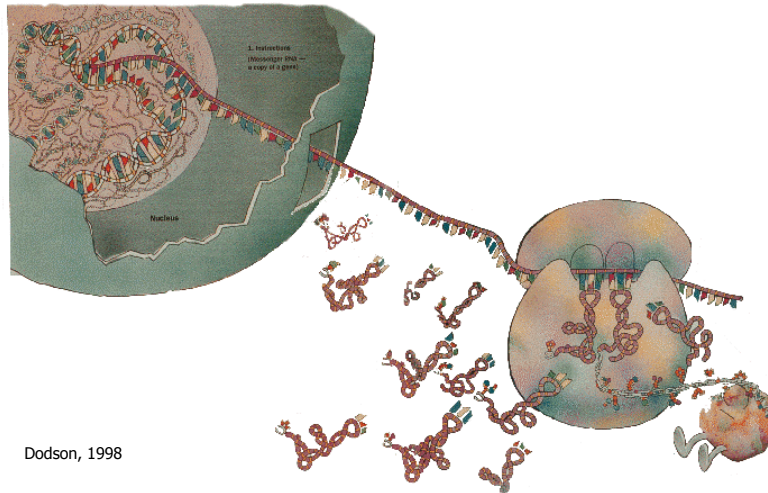
Z808006-01636.TIF

UCB Seminar

The Central Dogma of Molecular Biology



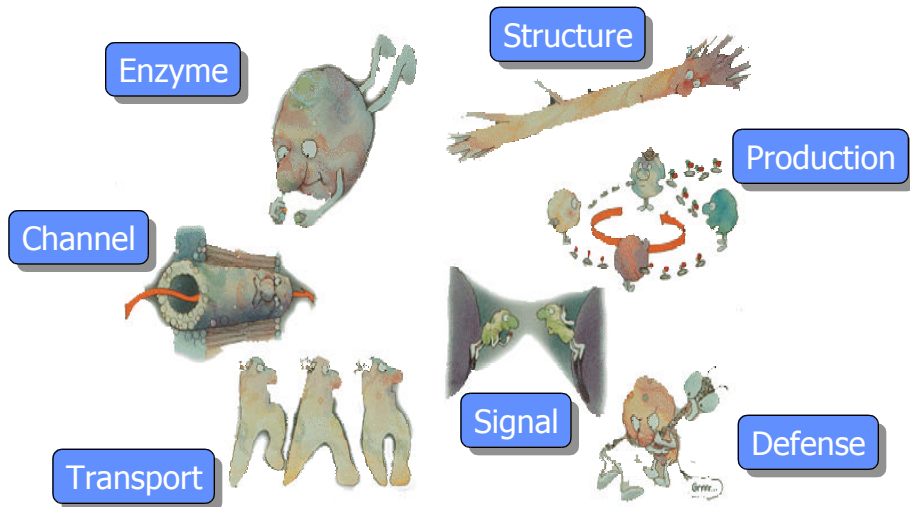
Protein Construction



Dodson, 1998

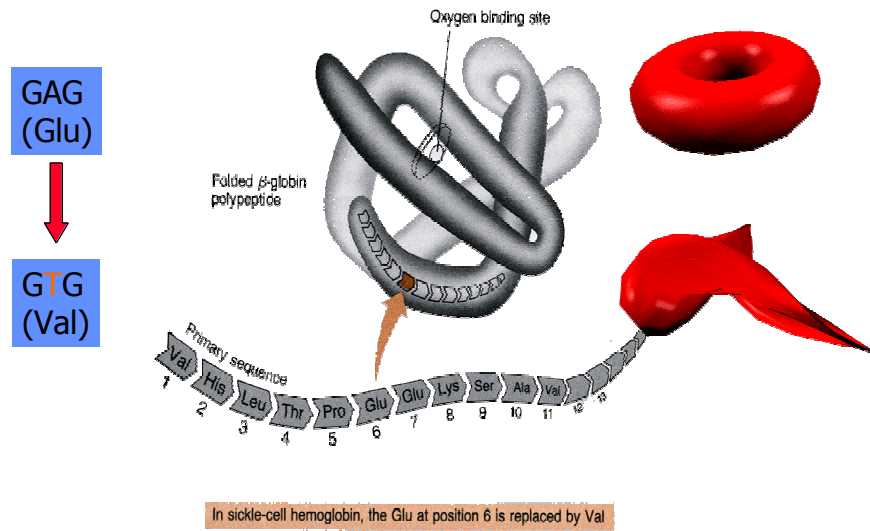
UCB Seminar

Protein Functions



UCB Seminar

Sickle Mutation



UCB Seminar

Is this a disease?

Normal

- disc-Shaped
- soft (like a bag of jelly)
- easily flow through small blood vessels
- lives for 120 days



Sickle





- sickle-Shaped
- hard (like a piece of wood)
- often get stuck in small blood vessels
- lives for 20 days or less



Well, it depends!

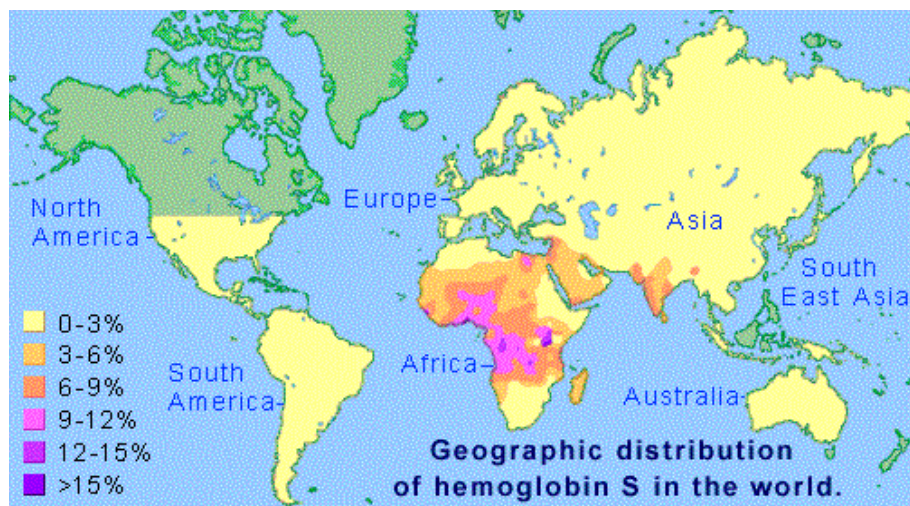
UCB Seminar

Genetic inheritance

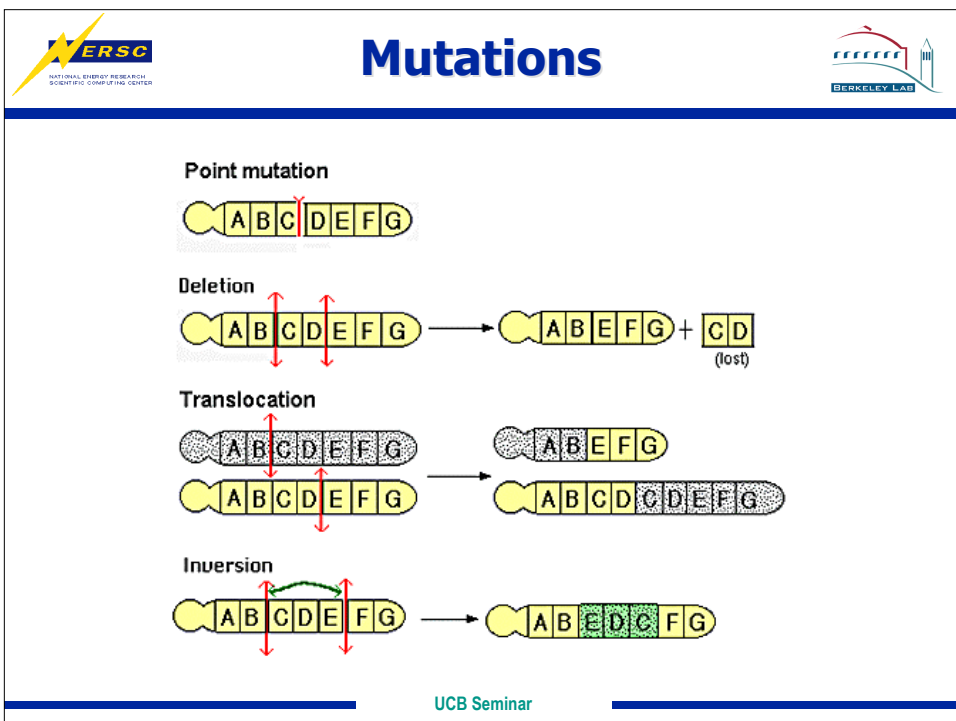
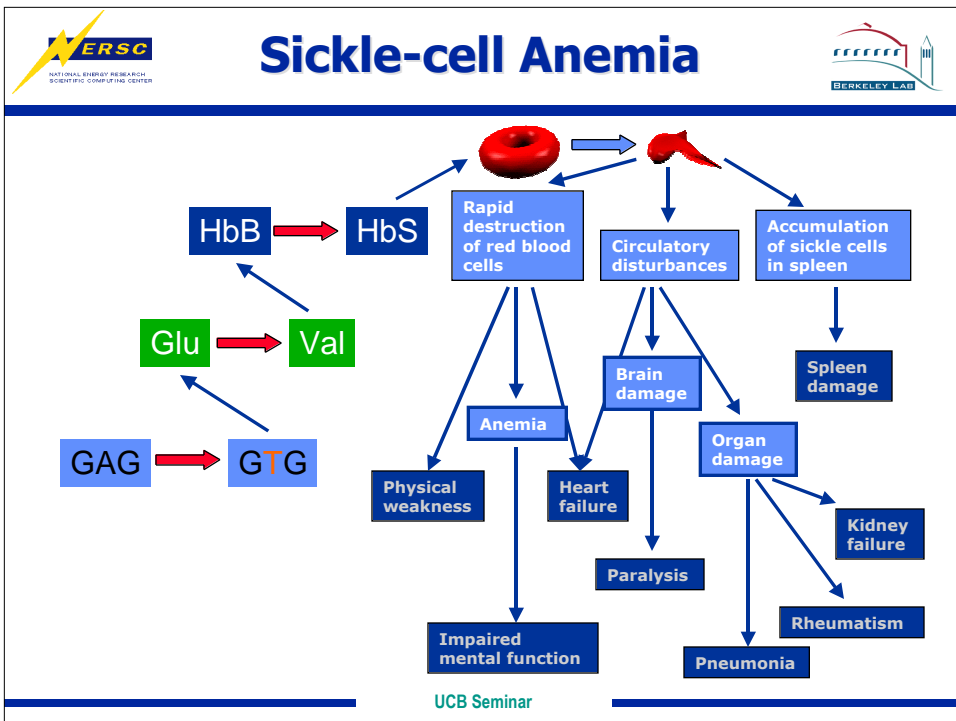
Normal/normal 	Sickle/normal  <ul style="list-style-type: none"> ■ reduced oxygen transport capability ■ normal life expectancy ■ resistant to malaria infection
Normal/sickle  <ul style="list-style-type: none"> ■ reduced oxygen transport capability ■ normal life expectancy ■ resistant to malaria infection 	Sickle/sickle  <ul style="list-style-type: none"> ■ impaired oxygen transport ■ reduced life expectancy

UCB Seminar

HbS Distribution



UCB Seminar





Small Changes BIG Effects



- DNA modifications can have a big effect
- Radiation causes DNA damage

US Department of Energy's
Long-standing program in radiation biology

UCB Seminar



Genome Project Timeline



- 1984
 - ✓ Department of Energy and Intl. Commission on Protection Against Environmental Mutagens and Carcinogens in Alta, Utah.
- 1986
 - ✓ DOE announces Human Genome Initiative
- 1987
 - ✓ NIH Director establishes Office of Genome Research
- 1988
 - ✓ NRC Mapping and Sequencing the Human Genome
 - ✓ Berkeley Lab launches Human Genome Center
- 1990 Human Genome I

UCB Seminar

Genome Timeline cont'd

- **September 1994**
 - ✓ First complete map of all human chromosomes one year ahead of schedule.
- **May 1995**
 - ✓ First genome sequenced: *H. influenzae*
- **May 1998**
 - ✓ Celera announces commercial project
 - ✓ Public effort regroups to five major centers
- **June 2000**
 - ✓ Joint announcement by NHGRI – Celera
- **February 2001**
 - ✓ Publication

We're done!

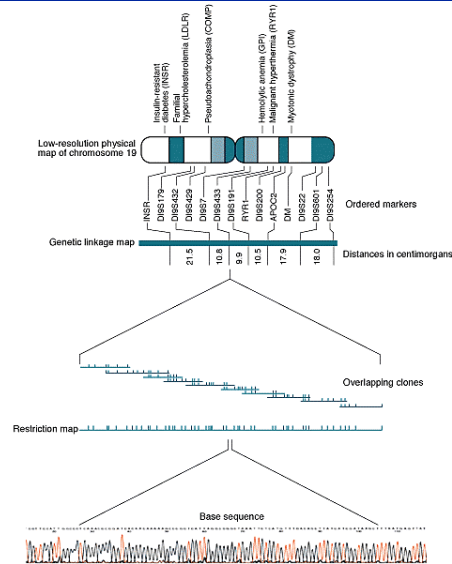
UCB Seminar

Genome Projects

1995	<i>H. influenzae</i>	2 Mb
1996	<i>S. cerevisiae</i>	12 Mb
1997	<i>E. coli</i>	5 Mb
1998	<i>C. elegans</i>	100 Mb
1999	Human Chromosome 22	34 Mb
2000	<i>D. melanogaster</i>	140 Mb
2000	<i>H. sapiens</i>	3,000 Mb

UCB Seminar

Mapping



UCB Seminar


DNA Sequencing

Read base code from storage medium!


- **Read length**
✓ about 550 bases per lane
- **Reader capacity**
✓ 96 lanes in parallel in about 2-4 hours
- **Sequencing lab**
✓ Dozens to hundreds of readers

UCB Seminar





DNA Analysis



Disassemble the base code!

- Find the genes
 - Heuristic signals
 - Inherent features
 - Intelligent methods

- Characterize each gene
 - Compare with other genes
 - Find functional components
 - Predict features

UCB Seminar

jlfkjdsfksdufpuurtwoljewkrnwensftarkthceresp
whireissnalitcommonsugpoirewnnbdaradisaccha
rhhjklooiuytideintoitsconstituqwertypoiuytrefvc
gentcompoundsglucoseanonhgstrefgsssdfructos
etransferprotsdvfgbnhnhyuimjuyheinintooutsid
eofchgfhgfhfhgellendhfdereasdfasdfldakldalskd
hbdbfsbfskdfiotueritupyuiklhfg;mkcdceihsuqtyw
qgwbsuagygvbnuwoyrpeqwurhdkagdiuqwtewo
ihrwererptergksfhakjdgadbsndfquiryoiweuterjgn
vsdvshsudihsuhfouwehosncodihtyeiqwtewre
oiewtujsdbbzvacdserdtgdhvjvibiunsgtdyfhnhjkojo
uywtreqqfbghjhkopwqoidhbrjgiiwhs

jlfkjdsfksdufpuurtwoljewkrnwensftarkthceresp
whireissnalitcommonsugpoirewnnbdaradisaccha
rhhjklooiuytideintoitsconstituqwertypoiuytrefvc
gentcompoundsglucoseanonhgstrefgsssdfructos
etransferprotsdvfgbnhnhyuimjuyheinintooutsid
eofchgfhgfhfhgellendhfdereasdfasdfldakldalskd
hbdbfsbfskdfiotueritupyuiklhfg;mkcdceihsuqtyw
qgwbsuagygvbnuwoyrpeqwurhdkagdiuqwtewo
ihrwererptergksfhakjdgadbsndfquiryoiweuterjgn
vsdvshsudihsuhfouwehosncodihtyeiqwtewre
oiewtujsdbbzvacdserdtgdhvjvibiunsgtdyfhnhjkojo
uywtreqqfbghjhkopwqoidhbrjgiiwhs

Gene Finding

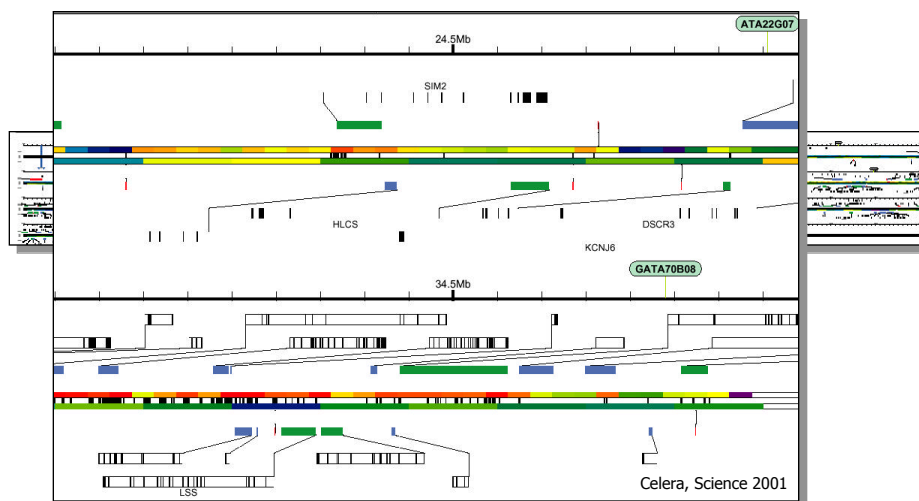
Start here.

**Split common sugar, a disaccharide, into
it's constituent compounds, glucose and
fructose. Transfer protein to outside of
cell.**

End here.

UCB Seminar

Celera Chromosome 21



UCB Seminar

- **2.91 Billion bases including gaps**
- **95% of the genome sequenced**
- **Quality about 99.95%**
- **30,000 genes**

- **40% genes with unknown function**
- **Single nucleotide changes not random and mostly outside of protein sequences**

- **The unit of inheritance**
- **A DNA sequence for a protein/RNA**
- **Fate**

Genes "FOR" a Condition

- **A rare occurrence in common observed conditions**
 - ✓ Cancer
 - ✓ Cystic fibrosis, Sickle cell anemia
- **A modifiable occurrence**
 - ✓ PKU, Diabetes, Hemochromatosis
- **A complex interaction with other genes/environment**
 - ✓ Predispositions to environmental factors
 - ✓ Dyslexia, Smoking

UCB Seminar

Cause and Effect

- **One gene \Rightarrow one phenotype**

Cystic Fibrosis

- 1/25 Americans are carriers
- 1/2500 live births (European ancestry)
- Almost 300 mutations with variable expression and population frequency

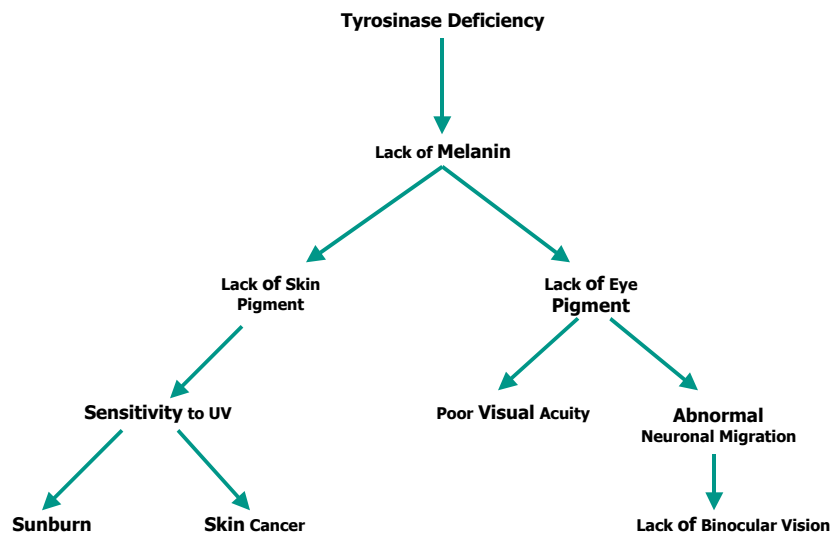
UCB Seminar

Cause and Effect

- One gene \Rightarrow one phenotype
✓ Cystic fibrosis
- One gene \Rightarrow many phenotypes

UCB Seminar

Effects resulting from Tyrosinase deficiency in albinism



UCB Seminar

Cause and Effect

- **One gene \Rightarrow one phenotype**
 - ✓ Cystic fibrosis
- **One gene \Rightarrow many phenotypes**
 - ✓ Tyrosinase deficiency, HbS
- **Many genes \Rightarrow one phenotype**

Diseases with Many Genes Involved

- **HEART DISEASE**
- **CANCER**
- **STROKE**
- **DEMENTIA**
- ...
- ...
- ...
- **BEHAVIOR**

3 Kinds of “Cancer-causing” Genes

- **ONCOGENES**
✓ ACCELERATOR ON
- **TUMOR-SUPPRESSOR GENES**
✓ BRAKES OFF
- **MUTATOR GENES**
✓ CONTROL THE RATE OF CHANGE IN OTHER GENES.

ALL CANCER IS GENETIC
BUT
NOT ALL CANCER IS
INHERITED.



Factoids On Colon Cancer



- **80% OF TUMORS ARE IN PEOPLE WITH NO FAMILY HISTORY**
- **A DISEASE OF AFFLUENCE**
- **USUALLY SLOW-GROWING TUMOR**
(1/2 inch polyp, 5 years, plus 5-10 years more to become malignant, plus 5-10 more years for symptoms)

UCB Seminar



Colon Cancer Genetics



- **FAP(Familial adenomatous polyposis-carpet of polyps)**
85% penetrance by 70, 1 gene
- **HNPCC(hereditary nonpolyposis colon cancer-fast growing, single lesion)**
50% penetrance, 4 genes

UCB Seminar

UCB Seminar





Life is the sum of ...



- Genetics
- Environment
- Education
- Opportunity
- ...

UCB Seminar



Nature and Nurture



- **Psychologists tend to emphasize the overwhelming importance of the environment in human development....**
- **Until they have their second child.**

UCB Seminar

- Genes predispose,
not predetermine.

**MOST DISEASES AND
BEHAVIORS ARE
INFLUENCED BY**

GENES.

**Information is the difference
that makes a difference.**

Are you better off knowing ?



ELSI: Ethical, Legal and Social Issues



- Education
- Intellectual property
- Privacy
- Discrimination
- Genetic testing
- Gene therapies

UCB Seminar



Questions



- What is a Genetic Disease?
- Should genetic status be protected in the eyes of the law, as sex, race, and national origin are?
- Excuse behavior based on genetic markers for mental disorders?
- Is there a right to genetic privacy?
- Limits of government involvement in genetic testing?
- How will we protect the disabled from genetic discrimination?

UCB Seminar



ELSI: A Few Good Issues



- **Legal Equality: What is normal?**
- **Pre-implantation Testing, and Embryo Selection**
- **Group vs. Individual Differences**
- **Coverage of Testing Costs**
- **Family Issues in Testing**
- **The Data Issues:**
 - ✓ **Employment**
 - ✓ **Forensics**
 - ✓ **Insurance**
 - ✓ **Reproductive Counseling**
 - ✓ **Others**

UCB Seminar



Types of Genetic Testing



- **Pre-symptomatic**
- **Diagnostic**
- **Pre-reproductive**
- **Neonatal**
- **Judicial**

UCB Seminar



Ethical Principles in Testing



- **Voluntariness**
- **Informed consent**
- **Confidentiality**
- **Privacy**
- **Equity**

UCB Seminar



Newborn Screening



Disorder	No. of States
<small>(incl. DC, Puerto Rico, Virgin Islands)</small>	
Phenylketonuria	52
Congenital hypothyroidism	52
Hemoglobinopathy	42
Galactosemia	38
Maple syrup urine disorder	22
Homocysteinuria	21
Biotinidase deficiency	14
Adrenal hyperplasia	8
Tyrosinemia	5
Cystic fibrosis	3

UCB Seminar

The PSA Test

- Early detection of prostate tumors
- Treatment: surgery or radiotherapy, both have risks
- Does it prolong life?

WHO KNOWS!!!

- Incidence of prostate cancer

Age	%
50	25
60	60
70	70
80	80+

- % of men dying of prostate cancer: 5%!

What Drives Testing?

- Profit
- Malpractice worries
- Curiosity
- Deep needs

Test Requirements

- Accuracy
- Sensitivity
- Specificity

Problems with Screening "Healthy" People

- False Positives
- False Negatives
- Mismatch between what is known and what can be done

Impact of False Positives and Negatives

- A screened population of 1 million
- A condition affecting 5% (50,000)
- A test that has a false-negative and a false-positive rate of 1%

- Result:
 - 500 of the 50,000 people who have the condition won't know it.
 - 9,500 people without the condition will think they do!

- WHAT ABOUT RETESTING?

UCB Seminar

Retesting

- Almost 20,000 people will have one positive and one negative!!
- 95 "poor" people who do **not** have the condition have tested +/+.
- This all costs money and anxiety!

UCB Seminar

Breast Cancer Gene Testing

- **BRCA1 implicated in 5-10% of breast cancers**
- **Many mutations with little idea of phenotype**

- **What is the correct response to having the gene?**

For breast cancer,
the risk is highest for
those who do **NOT** have
the gene!

Just as likely to get
the non-inherited form,
which accounts for
90% of the cases.

3 Questions

- What is the genome?
- Why is it important?
- What are the consequences?